PHIL 7001: Practice Questions for Final Exam

Professor Boris Babic

Basic Concepts of Classifications

1. Explain the concept of classification in machine learning to a beginner. Provide a simple real-world example to illustrate how classification works.
2. Describe the main difference between supervised and unsupervised learning in the context of classification.
3. List and briefly describe three real-world applications where classification is used for decision-making.
4. Consider a scenario where you want to predict the price of a house based on its features (e.g., size, number of bedrooms). Should you use regression or classification for this problem? Explain your choice.
5. How does the type of data (continuous vs. categorical) influence whether you should use regression or classification?

The Classification Models

1. In logistic regression, what does a probability value close to 1 signify, and what does a probability value close to 0 signify in the context of a binary classification problem?
2. What is the primary difference between a logistic regression and a linear regression model?
3. Interpret what a coefficient value in a logistic regression model tells you about the relationship between a feature and the outcome.
4. Explain why logistic regression models the log-odds of an event happening rather than modeling the probability directly. How does this choice benefit the modeling process?
5. Describe the shape of the sigmoid curve in logistic regression. What significance does this curve shape have in modeling probabilities?

6. Differentiate between linear regression and logistic regression. Highlight the key differences in terms of the problem type, response variable, and model output.
7. Define what "odds" mean in the context of logistic regression. Explain how the odds are transformed into probabilities using the sigmoid function.
8. Describe the purpose of the logit link function in logistic regression. How does it transform linear combinations of predictors into the log-odds?
9. Describe the key differences in interpreting model output between linear regression and logistic regression. How does the output of each model relate to the target variable?
10. Calculate the probability of a positive class given a logistic regression output (log-odds) of 2.5.
11. In SVM, data points that are closest to the hyperplane and influence its position are called _____.
12. You are given a dataset online_shopping that includes age, annual_income, and a binary variable repeat_customer (1 if the customer is a repeat customer, 0 otherwise).

```
online_shopping <- data.frame(
  age = c(22, 34, 45, 29, 39, 48, 31, 36, 27),
  annual_income = c(30, 45, 60, 35, 50, 70, 40, 55, 32),
  repeat_customer = c(0, 1, 1, 0, 1, 1, 0, 1, 0)
)
```

Fill in the blanks to create a random forest model predicting repeat_customer using the consumers' annual income and age information.

```
rf_model <- randomForest(__________ ~ _______ + _______, data = _______)
```

13. How would you modify the code if we now want to fit a logistic classifier instead?
14. In a classification problem, you have observed the following performance from a certain model. Calculate this model's performance metrics. Please round your answer to three decimal places.
    - True Positives (TP): 90
    - False Positives (FP): 20
    - True Negatives (TN): 65
    - False Negatives (FN): 5

    (a) Sensitivity

(b) Specificity

(c) Recall

(d) Precision

Neural Network and Deep Learning

1. What do Neural Language Models (NLMs) primarily approximate?
   (a) The distribution of words in a corpus
   (b) The probability function of word sequences
   (c) The syntactic structure of sentences
   (d) The frequency of word occurrence
2. Matching: Match the terms with their correct descriptions:
   (a) Language Model
   (b) Perplexity
   (c) Word2Vec
   - A measure of how well a model predicts a sample
   - A technique for word embedding
   - A probabilistic function over word sequences
3. What does a language model's probability function estimate?
   (a) The likelihood of a given word order in a sentence
   (b) The frequency of words in a large corpus
   (c) The similarity between different words
   (d) The number of possible sentences in a language
4. What is the purpose of calculating a weighted sum in each unit of a neural network?
   (a) To determine the output value of each neuron
   (b) To calculate the error rate of the model
   (c) To optimize the weights during backpropagation
   (d) To compare different neural network models
5. What does the softmax function accomplish in a neural network?

(a) It normalizes the output to a probability distribution.

(b) It reduces the complexity of the model.

(c) It speeds up the computation process.

(d) It increases the accuracy of the model.

6. According to the Universal Approximation Theorem, what can a neural network with at least one hidden layer approximate?

   (a) Any linear function

   (b) Any continuous function

   (c) Any discrete function

   (d) Any polynomial function

7. True or False: Stochastic Gradient Descent (SGD) is used in neural networks to iteratively adjust the model to minimize the loss.

8. True or False: The quality and quantity of data are not significant factors in the performance of deep learning models.

Reinforcement Learning, Natural Language Processing, and Large Language Models

1. What are the two main differences between Reinforcement Learning (RL) and classical Machine Learning (ML)?

   (a) RL uses labeled data; ML does not.

   (b) RL aims to make optimal sequential decisions; ML does not.

   (c) RL uses a reward function; ML uses a loss function.

   (d) RL's data is rewards of decisions; ML's data is labeled or non-labeled.

2. True or False: In reinforcement learning, an agent must exclusively focus on exploitation to achieve the best results.

3. What does a 2-dimensional Word2Vec visualization typically show?

   (a) The frequency of each word in the corpus.

   (b) The classification of words into categories.

   (c) Similar words being located nearby in space.

   (d) The sequence of words in a text.

4. *NLP and LLMs:* What is the primary difference between traditional NLP systems and Large Language Models (LLMs)?
   (a) LLMs use rule-based methods, whereas traditional NLP does not.
   (b) LLMs rely on handcrafted features, while traditional NLP uses deep learning.
   (c) LLMs can handle longer dependencies in text than traditional NLP systems.
   (d) Traditional NLP systems are capable of generating language, while LLMs are not.
5. *Transformer Architecture:* What was the key innovation introduced by the Transformer architecture in NLP?
   (a) The use of rule-based systems for language processing.
   (b) The introduction of self-attention mechanisms.
   (c) The reduction in the number of parameters in language models.
   (d) The shift from deep learning to statistical methods in NLP.
6. *Embedding in LLMs:* What does the embedding step in LLM processing do?
   (a) It reduces the dimensionality of the input data.
   (b) It maps tokens to vectors capturing identity and position.
   (c) It directly translates text into human language.
   (d) It predicts the next token in a sequence.
7. True or False: In LLMs, a token's new representation is a sum of all value vectors weighted by its attention scores towards other tokens.
8. Explain briefly the whole pipelines of LLMs. What steps does it involve?
9. What is the purpose of each step of the LLM pipeline? What are the input and output for each step?
10. What is the main purpose of Multi-Layer Perceptron(MLP) in the LLM pipeline?